

JC564 U.S. PTO
09/503757
02/14/00

PATENT OFFICE
JAPANESE GOVERNMENT

This is to certify that the annexed is a true copy of the following application as filed with this Office.

Date of Application: July 15, 1999

Application Number : P11-201988

Applicant(s) : KDD CORPORATION

December 10, 1999

Commissioner,
Patent Office

Takahiko KONDOU

H11-3085579

日 本 国 特 許 庁

PATENT OFFICE
JAPANESE GOVERNMENT



別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日

Date of Application:

1999年 7月15日

出 願 番 号

Application Number:

平成11年特許願第201988号

出 願 人

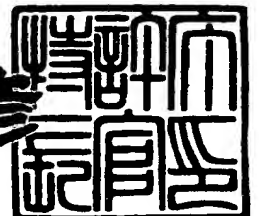
Applicant (s):

ケイディディ株式会社

1999年12月10日

特許庁長官
Commissioner,
Patent Office

近 藤 隆 彦



出証番号 出証特平11-3085579

【書類名】 特許願

【整理番号】 KDD-32

【提出日】 平成11年 7月15日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 17/00

【発明の名称】 情報自動フィルタリング方法および装置

【請求項の数】 4

【発明者】

【住所又は居所】 埼玉県上福岡市大原 2-1-15 株式会社ケイディデ
ィ研究所内

【氏名】 井ノ上 直己

【発明者】

【住所又は居所】 埼玉県上福岡市大原 2-1-15 株式会社ケイディデ
ィ研究所内

【氏名】 帆足 啓一郎

【発明者】

【住所又は居所】 埼玉県上福岡市大原 2-1-15 株式会社ケイディデ
ィ研究所内

【氏名】 橋本 和夫

【特許出願人】

【識別番号】 000001214

【氏名又は名称】 ケイディディ株式会社

【代理人】

【識別番号】 100083806

【弁理士】

【氏名又は名称】 三好 秀和

【電話番号】 03-3504-3075

【選任した代理人】

【識別番号】 100100712

【弁理士】

【氏名又は名称】 岩▲崎▼ 幸邦

【選任した代理人】

【識別番号】 100095500

【弁理士】

【氏名又は名称】 伊藤 正和

【選任した代理人】

【識別番号】 100101247

【弁理士】

【氏名又は名称】 高橋 俊一

【手数料の表示】

【予納台帳番号】 001982

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9506777

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 情報自動フィルタリング方法および装置

【特許請求の範囲】

【請求項 1】 インターネットを介して提供される情報のうち不適切情報を識別し、この識別した不適切情報の提供を阻止する情報自動フィルタリング方法であって、

提供の阻止を必要とする不適切な情報および提供の阻止を必要としない適切な情報を学習データとした自動学習により前記情報に含まれる単語に対して情報の提供を阻止する必要があるか否かを判定するために使用される単語の重みを求め

、

この求めた単語の重みを各単語に対応して重み付き単語リストとして記憶管理しておき、

インターネットを介して提供される情報を入力し、この情報に含まれる単語を抽出し、

この抽出した単語の各々に対する重みを前記重み付き単語リストから読み出し

、

この読み出した各単語の重みの総和を算出し、この算出した総和に基づき前記情報の提供を阻止すべきか否かを判定すること

を特徴とする情報自動フィルタリング方法。

【請求項 2】 前記単語の重みを求める処理は、前記不適切な情報と適切な情報に対してベクトル空間上で弁別できる線形識別関数に基づく自動学習により単語の重みを求めることを特徴とする請求項 1 記載の情報自動フィルタリング方法。

【請求項 3】 インターネットを介して提供される情報のうち不適切情報を識別し、この識別した不適切情報の提供を阻止する情報自動フィルタリング装置であって、

提供の阻止を必要とする不適切な情報および提供の阻止を必要としない適切な情報を学習データとした自動学習により前記情報に含まれる単語に対して情報の提供を阻止する必要があるか否かを判定するために使用される単語の重みを求め

る単語重み学習手段と、

この求めた単語の重みを各単語に対応して重み付き単語リストとして記憶管理する重み付き単語リスト格納手段と、

インターネットを介して提供される情報を入力する入力手段と、

この入力された情報に含まれる単語を抽出する単語抽出手段と、

この抽出した単語の各々に対する重みを前記重み付き単語リストから読み出し、この読み出した各単語の重みの総和を算出し、この算出した総和に基づき前記情報の提供を阻止すべきか否かを判定する判定手段と

を有することを特徴とする情報自動フィルタリング装置。

【請求項 4】 前記単語重み学習手段は、前記不適切な情報と適切な情報に対してベクトル空間上で弁別できる線形識別関数に基づく自動学習により単語の重みを求める手段を有することを特徴とする請求項 3 記載の情報自動フィルタリング装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、インターネットを介して提供されるメールを含む各種情報に対して、情報に出現する単語を抽出し、この単語に基づいて前記情報が不適切であるか否かを判定し、不適切な情報の提供を阻止する情報自動フィルタリング方法および装置に関する。

【0002】

【従来の技術】

インターネットの急速な広がりに伴い、限られた専門家の道具でしかなかったコンピュータはごく一般の家庭や学校などにも導入され始めている。このため、これまでコンピュータに触れることすらなかった多くの一般人でも気軽にインターネットにアクセスすることが可能になった。こうした背景の中、近年深刻な問題となっているのがインターネット上に氾濫するポルノ画像などの有害情報に対する子供のアクセスである。この問題に対処するため、アメリカでは政府機関がインターネット上の情報を検閲することを可能にした「通信品位法」という法律

が提案されたが、裁判の結果、表現の自由を保証する憲法に違反すると判決され、立法することができなかった。

【0003】

そこで最近注目されているのが「情報フィルタリング」という技術である。情報フィルタリングとは、ユーザがインターネット上の情報にアクセスする際にその情報の有害性をチェックし、有害と判定された場合は何らかの手段によりその情報へのアクセスをブロックするという技術である。

【0004】

現在市販されている有害情報フィルタリングソフトで取り入れられている手法は大きく以下の4つに分類される。

【0005】

- (1) 自己判定によるフィルタリング
- (2) 第三者の判定によるフィルタリング
- (3) 自動フィルタリング
- (4) 単語に対するスコア(得点)を利用する方式

ここではこの4つの手法について簡単に解説する。まず、自己判定によるフィルタリング手法ではWWW情報の提供者が自らのコンテンツの有害性について判定を行い、その結果をHTMLファイル内に記述する。フィルタリングソフトはこの記述された結果を参照し、有害と判断された場合にアクセスをブロックする。この手法によるフィルタリングを図4に示す。

【0006】

図4に示す自己判定に基づくフィルタリングでは、米国マサチューセッツ工科大学のWorld Wide Web Consortium が作成したPICS (Platform for Internet Content Selection) と呼ばれるインターネットコンテンツの評価を記述するための基準を使用している。PICSを使用することにより、コンテンツ提供者は簡単に自分の提供している情報を描写し、開示することができる。

【0007】

多くの場合、コンテンツ提供者がこのような評価結果を公開する際には、PICSによる評価結果を出力する評価機関のサービスを利用する。このような評価

機関の代表として、Recreational Software Advisory Council (R S A C) や SafeSurf といった団体があげられ、それぞれ独自に設定した基準による評価結果を提供している。コンテンツ提供者はこれらの機関からの評価結果を HTML ファイルのヘッダに記述する。図 5 にこの評価結果の記述例を示す。

【0008】

この自己判定はコンテンツ提供者の自主性に任せられるというのが現状である。そのため、多くのコンテンツ提供者がこの判定を受けようという意志を持たない限りは本手法による有効な有害情報フィルタリングは不可能であるといえる。

【0009】

次に第三者による判定に基づくフィルタリングについて説明する。有害情報フィルタリングソフトを作成している業者の中には、WWW 上のホームページの有害性を独自に判定し、その結果をフィルタリングソフトの判断基準とする手法を取り入れている。一般的には、この評価の結果として有害なホームページの URL 一覧が構築されている。この URL のリストはフィルタリングソフトと共にユーザに分配され、フィルタリングソフトの判断基準となる。多くの場合、フィルタリングソフトはこの有害 URL 一覧を定期的にダウンロードする仕組みになっている。第三者による判定に基づく有害情報フィルタリングの仕組みを図 6 に示す。

【0010】

このような仕組みを持つソフトウェアの代表的なものとして CyberPatrol があげられる。CyberPatrol は「暴力」「性行為」など 13 個のジャンルに対し、それぞれ有害 URL 一覧を持っており、これらのシステムに従って有害情報フィルタリングを行う。

【0011】

この手法で使用される有害 URL 一覧はそれぞれソフトウェア業者でホームページをアクセスし、判定を行うことによって作成・拡張されているため、新しく設立されたホームページや従来の URL から別の URL に移動したホームページには対処することは不可能である。従って、こうした評価対象外のページに対するフィルタリングには対処できないのが現状である。

【0012】

次に、自動フィルタリングについて説明する。有害情報フィルタリングソフトの中にはアクセスされたホームページの中身をチェックし、有害性の判断を行うものもある。

【0013】

具体的には、有害な情報、すなわち不適切な情報内に含まれるであろう単語を予め登録しておき、この登録した単語が情報内に出現するか否かをチェックし、前記登録した単語が含まれていた場合に情報の提供を阻止する方式である。例えば、ポルノ情報の提供を阻止する場合、情報内に”sex”や“xxx”といった文字列が含まれていた場合、その情報の提供を阻止する。この手法の応用として、登録した単語が情報内に含まれている割合が所定の閾値を上回った場合に情報の提供を阻止する方式もある。

【0014】

次に、単語に対するスコア（得点）を利用する方式について説明する。この方式は、不適切な情報内に含まれるであろう単語およびこの単語に対するスコアを予め登録しておき、この登録した単語が情報内に出現するか否かをチェックし、登録した単語が含まれていた場合に単語のスコアを合計し、この合計が所定の閾値を上回った場合に該情報の提供を阻止するものである。

【0015】

【発明が解決しようとする課題】

情報自動フィルタリングの大きな目的は不適切な情報を阻止する割合を増やすとともに、適切な情報が誤って阻止される割合を減らすことであるが、上述した各手法はそれぞれ一長一短があり、従来の情報自動フィルタリングでは十分なフィルタリング性能を得ることができないという問題がある。

【0016】

具体的には、従来の自動フィルタリング手法では、例えば”Sussex”というイギリスの町に関するホームページがブロックされるという悪例が報告されている。また、単語に対するスコアを利用する従来の方式では、単語および単語のスコアの設定がアドホックとなり、ユーザにとってどのように設定すれば最も有

効であるかに関して全く指針がなかった。そのため、提供を阻止すべき情報を阻止できなかったり、本来提供を阻止する必要のない情報が阻止されるなど、性能の点で問題があった。

【0017】

例えば、「女子高生」という単語は一般的にポルノ情報に頻出すると考え、「女子高生」という単語とそのスコアを40として登録したとする。その結果、「女子高生のサンプル画像、無料」という表現中に「女子高生」が含まれているため、この表現全体のスコアは40となる。また、同様に「女子高生の乗ったバスが北海道で事故」という表現についてもこの表現全体のスコアは40となり、これらの表現のスコアは同じになる。このため、閾値を20としたとすると、本来阻止する必要のない後者の表現が阻止されてしまうという問題があり、また閾値を50としたとすると、本来阻止すべき前者の表現が阻止されないという問題がある。これら2つの表現を区別するためには、「サンプル」「画像」「無料」などの単語や「バス」「北海道」「事故」といった単語にもスコアを設定する必要があることになるが、これらの単語は一般的にも良く利用される単語であり、スコアをどのように設定すれば良いかが明確でなく、スコアの設定により性能が大きく変動し、不適切な表現か否かの判定性能が十分に得られないという問題がある。

【0018】

本発明は、上記に鑑みてなされたもので、その目的とするところは、単語の重みを容易かつ適確に設定し、この単語の重みを利用して情報が不適切であるか否かを適確に判定する情報自動フィルタリング方法および装置を提供することにある。

【0019】

【課題を解決するための手段】

上記目的を達成するため、請求項1記載の本発明は、インターネットを介して提供される情報のうち不適切情報を識別し、この識別した不適切情報の提供を阻止する情報自動フィルタリング方法であって、提供の阻止を必要とする不適切な情報および提供の阻止を必要としない適切な情報を学習データとした自動学習に

より前記情報に含まれる単語に対して情報の提供を阻止する必要があるか否かを判定するために使用される単語の重みを求め、この求めた単語の重みを各単語に対応して重み付き単語リストとして記憶管理しておき、インターネットを介して提供される情報を入力し、この情報に含まれる単語を抽出し、この抽出した単語の各々に対する重みを前記重み付き単語リストから読み出し、この読み出した各単語の重みの総和を算出し、この算出した総和に基づき前記情報の提供を阻止すべきか否かを判定することを要旨とする。

【0020】

請求項1記載の本発明にあっては、提供の阻止を必要とする不適切な情報および提供の阻止を必要としない適切な情報を学習データとした自動学習により単語の重みを求め、この求めた単語の重みを各単語に対応して重み付き単語リストとして記憶管理しておき、インターネットを介して提供される情報に含まれる単語を抽出し、この抽出した単語の各々に対する重みを重み付き単語リストから読み出し、この読み出した各単語の重みの総和を算出し、この総和に基づき前記情報の提供を阻止すべきか否かを判定するため、従来アドホックに設定しなければならなかった単語の重みを自動学習により適確に求め、この適確に求めた単語の重みを利用して情報が不適切な情報であるか否かを適確に判定し、不適切な情報の提供を阻止することができる。

【0021】

また、請求項2記載の本発明は、請求項1記載の発明において、前記単語の重みを求める処理が、前記不適切な文書と適切な文書に対してベクトル空間上で弁別できる線形識別関数に基づく自動学習により単語の重みを求めることを要旨とする。

【0022】

請求項2記載の本発明にあっては、不適切な文書と適切な文書に対してベクトル空間上で弁別できる線形識別関数に基づく自動学習により単語の重みを求めるため、単語の重みを適確に設定することができる。

【0023】

更に、請求項3記載の本発明は、インターネットを介して提供される情報のう

ち不適切情報を識別し、この識別した不適切情報の提供を阻止する情報自動フィルタリング装置であって、提供の阻止を必要とする不適切な文書および提供の阻止を必要としない適切な文書を学習データとした自動学習により前記文書に含まれる単語に対して情報の提供を阻止する必要があるか否かを判定するために使用される単語の重みを求める単語重み学習手段と、この求めた単語の重みを各単語に対応して重み付き単語リストとして記憶管理する重み付き単語リスト格納手段と、インターネットを介して提供される情報を入力する入力手段と、この入力された情報に含まれる単語を抽出する単語抽出手段と、この抽出した単語の各々に対する重みを前記重み付き単語リストから読み出し、この読み出した各単語の重みの総和を算出し、この算出した総和に基づき前記情報の提供を阻止すべきか否かを判定する判定手段とを有することを要旨とする。

【0024】

請求項3記載の本発明にあっては、提供の阻止を必要とする不適切な文書および提供の阻止を必要としない適切な文書を学習データとした自動学習により単語の重みを求め、この求めた単語の重みを各単語に対応して重み付き単語リストとして記憶管理しておき、インターネットを介して提供される情報に含まれる単語を抽出し、この抽出した単語の各々に対する重みを重み付き単語リストから読み出し、この読み出した各単語の重みの総和を算出し、この総和に基づき前記情報の提供を阻止すべきか否かを判定するため、従来アドホックに設定しなければならなかった単語の重みを自動学習により適確に求め、この適確に求めた単語の重みを利用して情報が不適切な情報であるか否かを適確に判定し、不適切な情報の提供を阻止することができる。

【0025】

請求項4記載の本発明は、請求項3記載の発明において、前記単語重み学習手段が、前記不適切な文書と適切な文書に対してベクトル空間上で弁別できる線形識別関数に基づく自動学習により単語の重みを求める手段を有することを要旨とする。

【0026】

請求項4記載の本発明にあっては、不適切な文書と適切な文書に対してベクト

ル空間上で弁別できる線形識別関数に基づく自動学習により単語の重みを求めるため、単語の重みを適確に設定することができる。

【0027】

【発明の実施の形態】

次に、図1を参照して、本発明の実施形態に係る情報自動フィルタリング装置について説明する。同図に示す情報自動フィルタリング装置は、単語の重みを自動学習により求め、この自動学習で求めた単語の重みを利用して情報が不適切であるか否かを判定し、不適切な情報の提供を阻止するものであり、インターネットを介して提供されるHTML情報を入力する入力部1、この入力部1を介して入力された情報に出現する単語抽出部3、提供の阻止を必要とする不適切な情報である文書および提供の阻止を必要としない適切な情報である文書を学習データとした自動学習により前記文書に含まれる単語に対して情報の提供を阻止する必要があるか否かを判定するために使用される単語の重みを求める重み付き単語リスト学習部60、この重み付き単語リスト学習部60で求めた単語の重みを各単語に対応して重み付き単語リストとして記憶管理する重み付き単語リスト格納部50、単語抽出部3で抽出された単語および該単語に対して重み付き単語リスト格納部50から得られた単語の重み w に基づき入力部1から入力された情報の提供を阻止すべきか否かを判定する自動フィルタリング部30、および該自動フィルタリング部30で得られた判定結果を出力する出力部40から構成されている。

【0028】

本実施形態の情報自動フィルタリング装置は、重み付き単語リスト学習部60において単語の重みを自動学習により予め取得し、この自動学習で得た単語の重みを利用することを特徴とする。この単語の重みの自動学習の方法を図2のフローチャートに示す単語重みの学習アルゴリズムで行われるものである。すなわち、図2に示す学習アルゴリズムでは、学習データの集合 $E = \{d_1, \dots, d_n\}$ として提供の阻止を必要とする不適切な情報および提供の阻止を必要としない適切な情報を重み付き単語リスト学習部60に入力し、この入力された不適切な情報と適切な情報をベクトル空間上で弁別する線形識別関数から単語の重みを取得

する。具体的には次ぎの手順で行う。

【0029】

まず、入力部 1 から入力された HTML 文書をベクトル空間モデルによって表現する。すなわち、すべての文書を表現する n 個の単語を選択し、それぞれの文書を n 次元のベクトルで次式のように表現する。

【0030】

【数 1】

$$\vec{V}_d = (f_{d1}, \dots, f_{di}, \dots, f_{dn}) \quad \dots (1)$$

このベクトルの各要素は、各々単語の文書 d での出現頻度を正規化したものである。単語の出現頻度の正規化には次に示す数式で表される $TF * IDF$ という手法を用いている。

【0031】

【数 2】

$$f_{di} = tf_{di} * \log \left[\frac{N}{df_i} \right] \quad \dots (2)$$

ここで、 tf_{di} は単語 i が文書 d に出現する頻度、 N はすべての文書の数、 df_i は単語 i が出現する文書の数である。

【0032】

自動フィルタリングは、次に示す数式で表される線形識別関数によって行われ、この関数によって単語の重みの総和 $Dis(d)$ が計算される。

【0033】

【数 3】

$$Dis(d) = \sum_{i=1}^n w_i * f_{di} \quad \dots (3)$$

ここで、 w_i は各単語 i に対する重みであり、 f_{di} は上式 (2) の値であり、文書における各単語の f_{di} 値である。

【0034】

上述した式(3)から、総和 $Dis(d)$ が0より大きい場合、前記文書は有害であり、0以下である場合、無害であると判定される。

【0035】

なお、上述した各単語 i に対する重みは文書 d が有害な場合、総和 $Dis(d) > 0$ となり、無害な場合、総和 $Dis(d) \leq 0$ となるように設定される。

【0036】

次に、この単語の重みの学習アルゴリズムについて図2に示すフローチャートを参照して説明する。なお、この単語の重みの学習には **perceptron learning algorithm (PLA)** を使用している。

【0037】

図2においては、まず各種パラメータを設定する(ステップS51)。このパラメータとしては、各単語の重みの集合 $W = (w_1, \dots, w_n)$ 、 N 個の学習データ $E = \{d_1, \dots, d_n\}$ 、定数 η 、最大学習回数 Max 、図2に示す学習処理を繰り返し行う学習回数 m がある。

【0038】

それから、全ての文書を表現する単語のうち頻度の高い n 個の単語を選択する(ステップS52)。

【0039】

次に、単語の重みの集合 W を初期化する(ステップS53)。この初期化では、各単語の重みに乱数を入力する。それから、すべての学習データに対して前記単語重みの総和 $Dis(d)$ を上式(3)により計算する(ステップS55)。

【0040】

そして、この計算の結果、すべての無害な文書 d について総和 $Dis(d) \leq 0$ であり、かつすべての有害な文書 d について総和 $Dis(d) > 0$ であるか否かをチェックし(ステップS57)、そうである場合には、処理を終了するが、そうでない場合には、このように誤って分類されたすべての文書 d について次のステップS61、S63で示すように重みの変化度合 S を補正する(ステップS59)。

【0041】

すなわち、ステップ S61 では、文書 d_i が有害であって、かつ総和 $Dis(d) \leq 0$ の場合には、重み変化度合 S を増加するように補正し、またステップ S63 では、文書 d_i が無害であって、かつ総和 $Dis(d) > 0$ の場合には、重み変化度合 S を低減するように補正する。

【0042】

そして、このように補正された重み変化度合 S を使用して単語重みの集合 W をステップ S65 で示す式のように補正する。それから、学習回数 m を +1 インクリメントし（ステップ S67）、この学習回数 m が最大学習回数 Max より小さいか否かをチェックし（ステップ S69）、また最大学習回数 Max より小さい場合には、ステップ S55 に戻り、ステップ S57 に示した条件が満たされるまで、ステップ S55 以降の処理を繰り返し行う。そして、最終的に n 個の単語に対する単語重みの集合が求まる。

【0043】

重み付き単語リスト学習部 60 で取得された各単語の重みは、各単語に対応して重み付き単語リストとして重み付き単語リスト格納部 50 に格納される。次に示す表 7 は、重み付き単語リスト格納部 50 に格納されている重み付き単語リストを示す表であり、各単語に対応して単語重み w が格納されている。

【0044】

【表 1】

重み付き単語リスト	
単語	単語重み w
⋮	
画 像	10.9
⋮	⋮
サンプル	18.7
⋮	⋮
事 故	-16.6
⋮	⋮
女子高生	82.2
⋮	⋮
バ ス	-101.9
⋮	⋮
北 海 道	-112.5
⋮	⋮
無 料	-6.3
⋮	⋮

次に、このように重み付き単語リスト学習部 60 で得られ、重み付き単語リスト格納部 50 に格納された単語重みに基づきインターネットから提供された情報が不適切な情報であるか否かを判定する処理について説明する。

【0045】

図 1 において、入力部 1 から入力されたインターネットからの情報は、単語抽出部 3 で、重み付き単語リスト格納部 50 に格納されている単語リストと照合し、入力情報中に出現する単語とその出現頻度を求める。また、同時に出現した単語の重み w も重み付き単語リスト格納部 50 から求め、出現単語とその頻度および重みを自動フィルタリング部 30 に供給する。自動フィルタリング部 30 は、この入力された単語に対する重み w と出現頻度から、入力情報中に出現した全て

の単語に対する重み w の総和を算出し、この総和を所定の閾値と比較し、総和が閾値よりも大きい場合不適切な情報と判定し、総和が閾値よりも小さい場合、適切な情報と判定し、この判定結果を出力部 40 から出力する。

【0046】

具体的に説明する。表 1 に示すように、重み付き単語リスト学習部 60 では、予め入力された学習データから「画像」の重みは 10.9、「サンプル」の重みは 18.7、「事故」の重みは -16.6、「女子高生」の重みは 82.2、「バス」の重みは -101.9、「北海道」の重みは -112.5、「無料」の重みは -6.3 と求まり、重み付き単語リスト格納部 50 に格納しているので、この結果を利用すると、例えば「女子高生の乗ったバスが北海道で事故」という表現全体に対しては、自動フィルタリング部 30 で各単語の重みの総和を求め、 $82.2 - 101.9 - 112.5 - 16.6 = -148.8$ となる。また、「女子高生のサンプル画像、無料」の表現全体に対しては、自動フィルタリング部 30 で各単語の総和を求め、 $82.2 + 18.7 + 10.9 - 6.3 = 105.5$ となる。そして、図 2 の処理と同様に閾値を 0 とすると、「女子高生の乗ったバスが北海道で事故」という表現は閾値を下回るので、情報の提供は阻止されず、また「女子高生のサンプル画像、無料」という表現は閾値を上回るので、情報の提供は阻止されるというように正しく判定することができる。

【0047】

次に、図 4 および図 6 を参照して、本発明の他の実施形態に係る自動フィルタリング装置について説明する。図 4 に示す自動フィルタリング装置は、図 6 で説明した学習により単語リストを作成する情報自動フィルタリング装置 25 に対して第三者判定フィルタリング処理部 23 および該第三者判定フィルタリング処理部 23 で有害 URL を参照するために使用される有害 URL 一覧テーブル格納部 17 が付加されている。

【0048】

有害 URL 一覧テーブル格納部 17 は、有害情報を提供する URL を有害 URL 一覧テーブルとして格納しているものであり、第三者判定フィルタリング処理部 23 は、前記入力部 1 から入力された HTML 文書の URL を有害 URL 一覧

テーブル格納部 17 の有害 URL 一覧テーブルに登録されている各 URL と照合し、一致する URL があるか否かを判定するものである。

【0049】

図 6 は、図 4 に示す自動フィルタリング装置の更に詳細な構成を示すブロック図である。図 6 に示す自動フィルタリング装置は、図 6 に示した学習により作成した重み付き単語リストを用いた情報自動フィルタリング装置を構成する入力部 1、単語抽出部 3、重み付き単語リスト格納部 50、自動フィルタリング部 30、出力部 40 に加えて、図 4 の第三者判定フィルタリング処理部 23 に対応する URL リストに基づくフィルタリング部 15 および有害 URL 一覧テーブル格納部 17 を有している。

【0050】

このように構成される自動フィルタリング装置、すなわち第三者判定フィルタリング処理部による URL リスト一覧と学習により作成した重み付き単語リストを用いた情報自動フィルタリング装置によるフィルタリング処理では、まずインターネット 21 を介して入力された HTML 文書は、その URL が有害 URL 一覧テーブル格納部 17 の有害 URL 一覧テーブルに登録されている各 URL と照合され、一致する URL があるか否かが判定される。そして、有害 URL 一覧テーブル格納部 17 の有害 URL 一覧テーブルに登録された URL と一致する場合には、この URL が示す情報の提示は阻止される。

【0051】

URL リストに基づくフィルタリング部 15 による有害 URL 一覧テーブルを参照した判定の結果、有害 URL 一覧テーブル格納部 17 の有害 URL 一覧テーブルに登録されている URL と一致するものがない場合には、学習により作成した重み付き単語リストを用いた情報自動フィルタリング装置 25 によるフィルタリングが図 6 で説明したように行われる。

【0052】

このように本実施形態では、第三者による判定に基づくフィルタリングと学習により作成した重み付き単語リストを用いたフィルタリングの両方が行われるため、有害情報を適確に検出して阻止することができる。

【0053】

【発明の効果】

以上説明したように、本発明によれば、提供の阻止を必要とする不適切な情報および提供の阻止を必要としない適切な情報を学習データとした自動学習により単語の重みを求め、この単語の重みを各単語に対応して重み付き単語リストとして記憶管理し、インターネットを介して提供される情報に含まれる単語を抽出し、この抽出した単語の各々に対する重みを重み付き単語リストから読み出し、各単語の重みの総和を算出し、この総和に基づき情報の提供を阻止すべきか否かを判定するので、従来アドホックに設定しなければならなかった単語の重みを自動学習により適確に求め、この適確に求めた単語の重みを利用して情報が不適切な情報であるか否かを適確に高い性能で判定し、不適切な情報の提供を阻止することができる。

【図面の簡単な説明】

【図1】

本発明の別の実施形態に係る情報自動フィルタリング装置の構成を示すブロック図である。

【図2】

図1に示すフローチャートに使用されている単語重みの設定手順を示すフローチャートである。

【図3】

本発明の他の実施形態に係る自動フィルタリング装置の概要構成を示す説明図である。

【図4】

従来の自己判定に基づくフィルタリングを説明するための図である。

【図5】

図4に示した自己判定に基づくフィルタリングの一例としてRSACi とSafeSurf による評価結果の記述例を示す図である。

【図6】

従来の第三者による判定に基づく有害情報フィルタリングを説明するための図

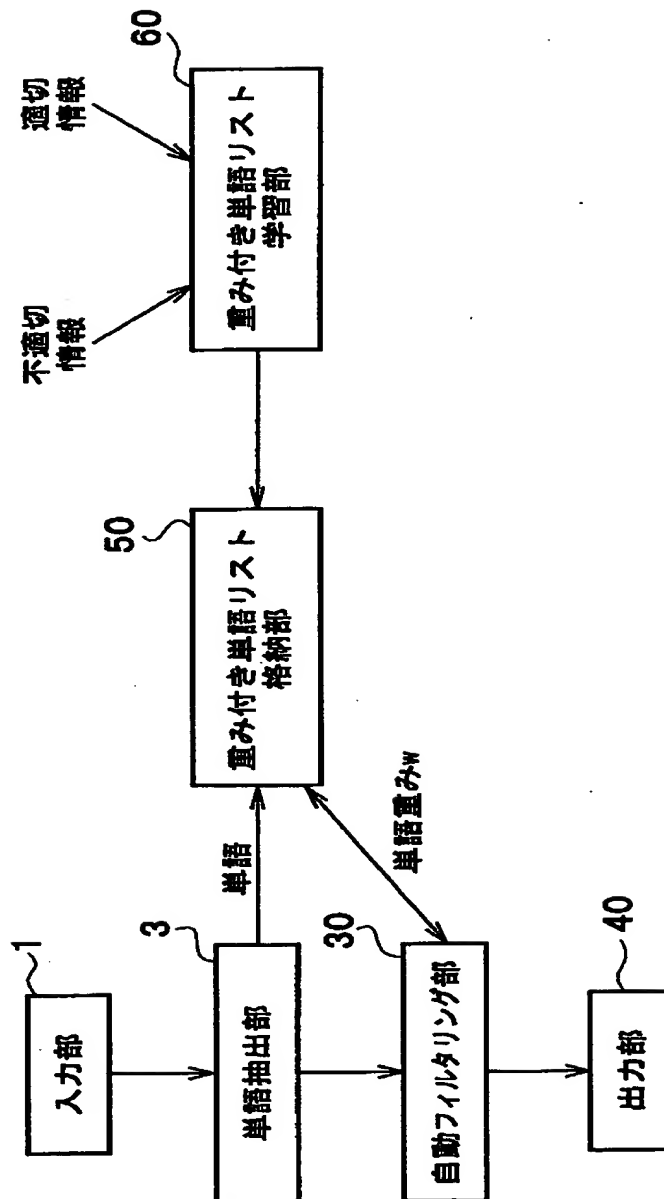
である。

【符号の説明】

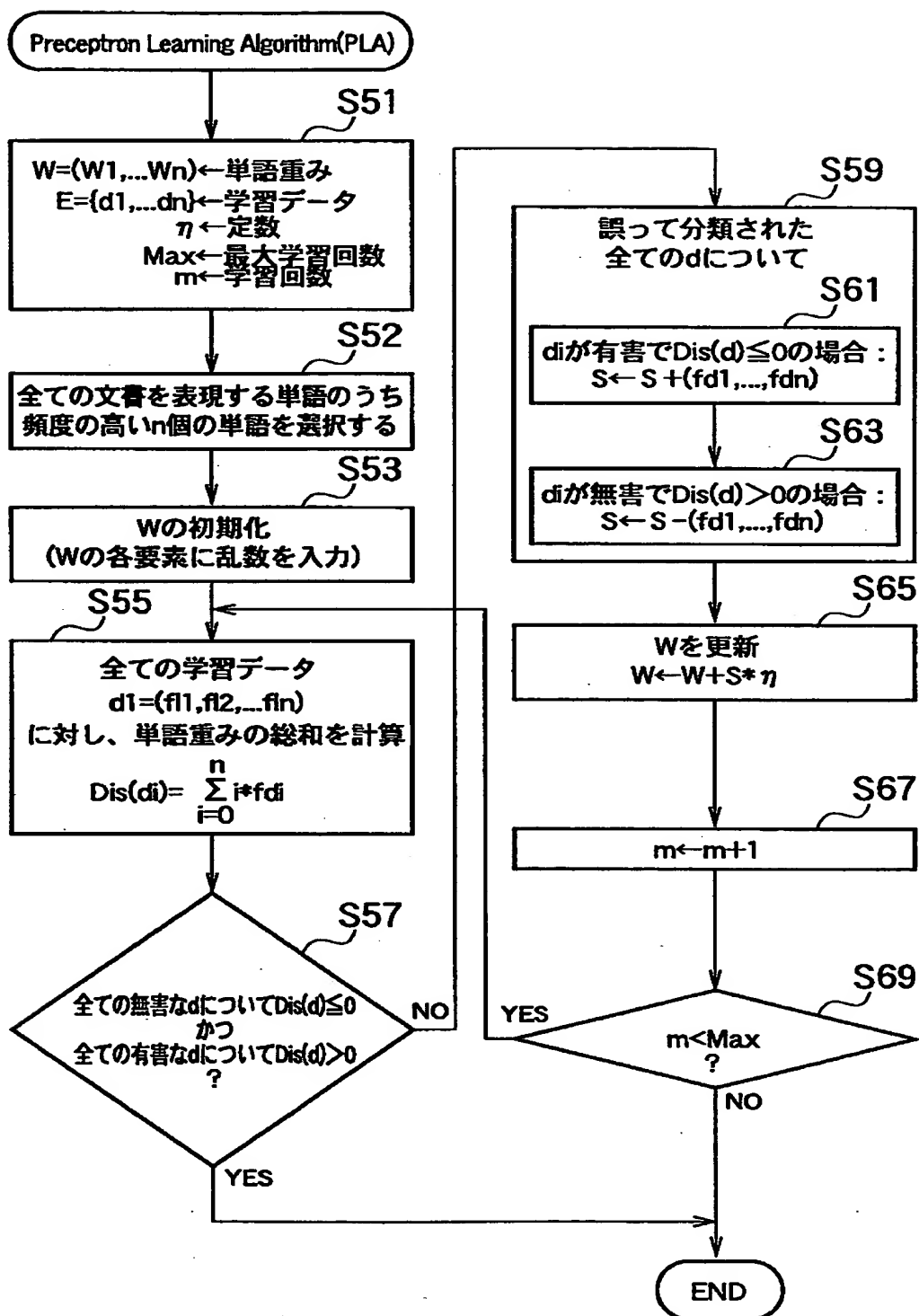
- 1 入力部
- 3 単語抽出部
- 30 自動フィルタリング部
- 50 重み付き単語リスト格納部
- 60 重み付き単語リスト学習部

【書類名】 図面

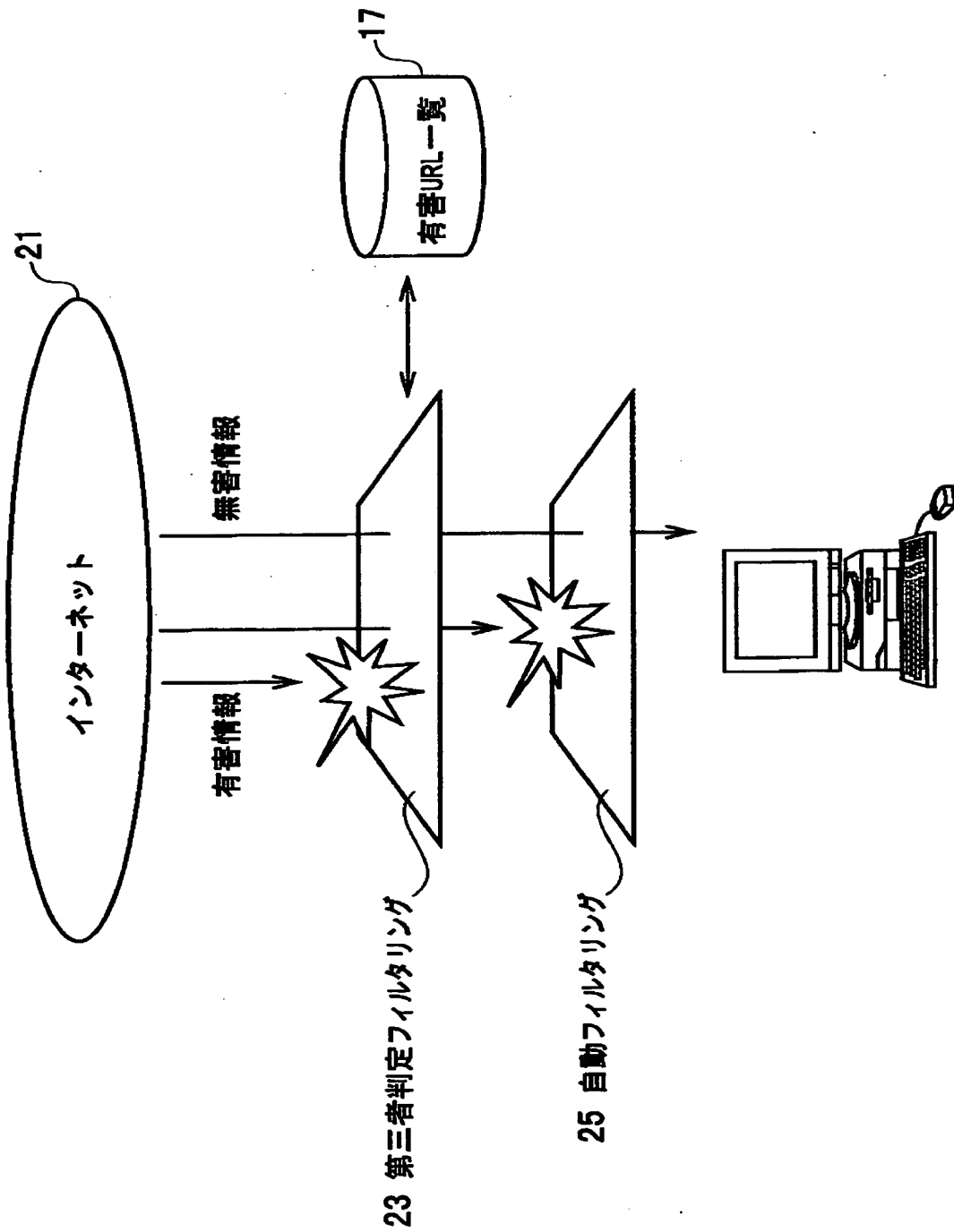
【図 1】



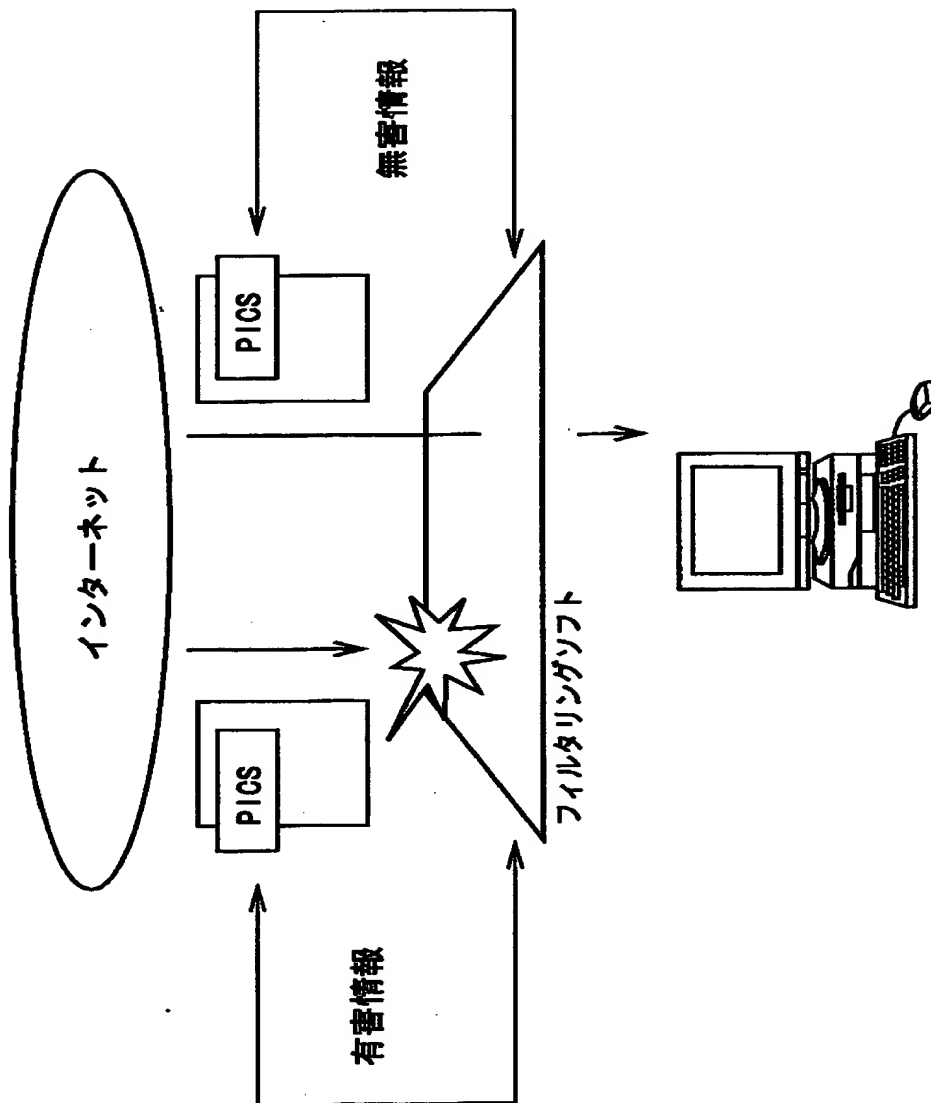
【図 2】



【図 3】



【図 4】



【図 5】

```

<HTML>
<HEAD>
<TITLE>
Example of RSAC1 and SafeSurf Rated Page
</TITLE>

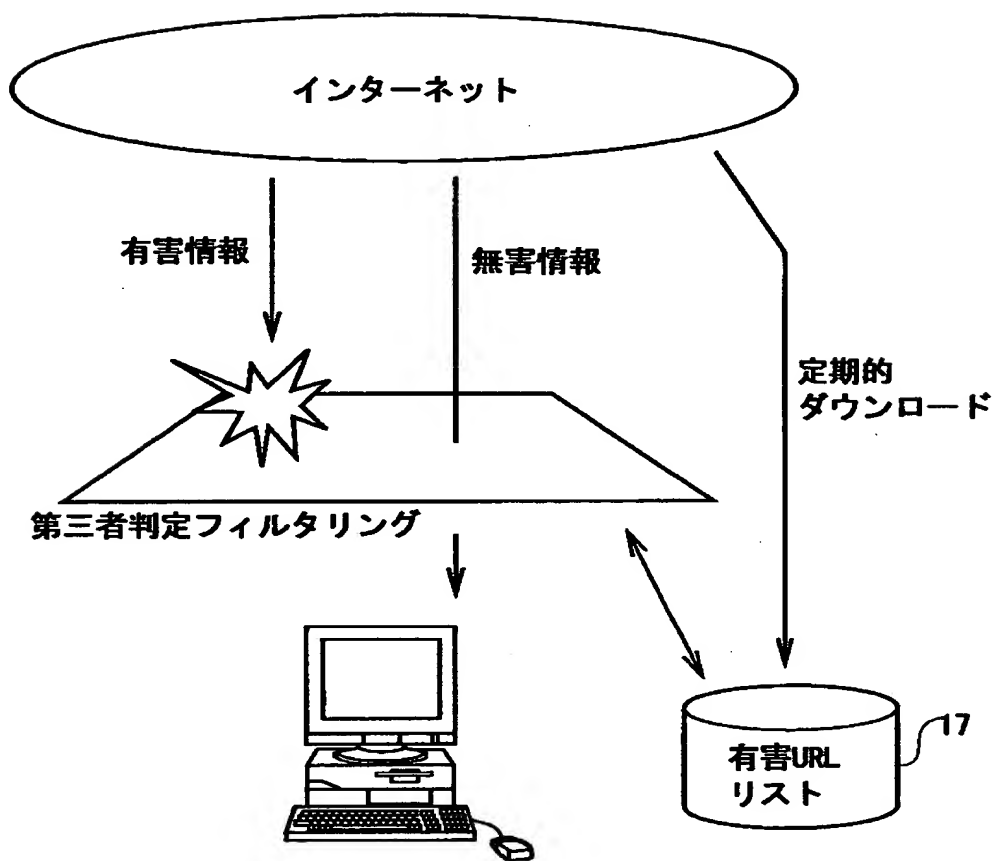
<META http-equiv="PICS-Label" content- '(PICS-1.0
"http://www.rsac.org/ratingav01.html"
1 gen true comment "RASC1 North Amerika Serber" by
"marks@icrosys.com" for "http://www.samplesite.com" on
"1996.09.05T02:15-0800" exp "1997.09.05T02:15-0800"
r (n 2 s 0 v 1 1 2))'>

<META http-equiv="PICS-Label" content- '(PICS-1.1
"http://www.classify.org/safesurf/" 1 gen t for
"http://www.samplesite.com/"
r(SS--000 2 SS--100 1))'>

</HEAD>
<BODY>
...
```

RSACiとSafeSurfによる評価結果の記述例

【図 6】



【書類名】 要約書

【要約】

【課題】 単語の重みを容易かつ適確に設定し、この単語の重みを利用して情報が不適切であるか否かを適確に判定する情報自動フィルタリング方法および装置を提供する。

【解決手段】 重み付き単語リスト学習部 60 に学習データとして提供の阻止を必要とする不適切な情報と提供の阻止を必要としない適切な情報を入力し、不適切な情報と適切な情報をベクトル空間上で弁別する線形識別関数から単語の重みを取得して重み付き単語リストとして重み付き単語リスト格納部 50 に格納し、入力部 1 からの情報から単語抽出部 3 で単語を抽出し、この単語の重み w を重み付き単語リスト格納部 50 から取得して自動フィルタリング部 30 に入力し、これらの単語の重み w の総和を算出し、総和が閾値よりも大きい場合不適切な情報と判定し、総和が閾値よりも小さい場合、適切な情報と判定する。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [000001214]

1. 変更年月日	1998年12月 3日
[変更理由]	名称変更
住 所	東京都新宿区西新宿2丁目3番2号
氏 名	ケイディディ株式会社